DOCUMENT RESUME

ED 397 087 TM 025 121

AUTHOR Mislevy, Robert J.; Gitomer, Drew H.
TITLE The Role of Probability-Based Inference in an Intelligent Tutoring System.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO ETS-RR-95-42-ONR
PUB DATE Oct 95
CONTRACT N00014-91-J-4101
NOTE 51p.; Additional support from the Armstrong Laboratories of the United States Air Force and the National Center for Research on Evaluation, Standards, and Student Testing.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Clinical Diagnosis; *Educational Theories; Hydraulics; *Intelligent Tutoring Systems; Learning; Networks; *Probability; *Statistical Inference
IDENTIFIERS *HYDRIVE

ABSTRACT

Probability-based inference in complex networks of interdependent variables is an active topic in statistical research, spurred by such diverse applications as forecasting, pedigree analysis, troubleshooting, and medical diagnosis. This paper concerns the role of Bayesian inference networks for updating student models in intelligent tutoring systems (ITSs). Basic concepts of the approach are briefly reviewed, but the emphasis is on the considerations that arise when one attempts to operationalize the abstract framework of probability-based reasoning in a practical ITS context. The discussion revolves around HYDRIVE, an ITS for learning to trouble-shoot an aircraft hydraulics system. HYDRIVE supports generalized claims about aspects of student proficiency through probability-based combination of rule-based evaluations of specific actions. The paper highlights the interplay among inferential issues, the psychology of learning in the domain, and the instructional approach upon which the ITS is based. (Contains 2 tables, 9 figures, and 37 references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 087

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

RR-95-42-ONR

THE ROLE OF PROBABILITY-BASED INFERENCE IN AN INTELLIGENT TUTORING SYSTEM

Robert J. Mislevy
Drew H. Gitomer

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-91-J-4101
R&T 4421573-01

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, NJ

October 1995

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

BEST COPY AVAILABLE

TM025721

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	Final

4. TITLE AND SUBTITLE
The Role of Probability-Based Inference in an Intelligent Tutoring System

5. FUNDING NUMBERS
G. N00014-91-J-4101
PE. 61153N
PR. RR 04204
TA. RR 04204-01
WU. R & T 4421537-01

6. AUTHOR(S)
Robert J. Mislevy and Drew H. Gitomer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-95-42-ONR

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*

Probability-based inference in complex networks of interdependent variables is an active topic in statistical research, spurred by such diverse applications as forecasting, pedigree analysis, troubleshooting, and medical diagnosis. This paper concerns the role of Bayesian inference networks for updating student models in intelligent tutoring systems (ITSs). Basic concepts of the approach are briefly reviewed, but the emphasis is on the considerations that arise when one attempts to operationalize the abstract framework of probability-based reasoning in a practical ITS context. The discussion revolves around HYDRIVE, an ITS for learning to troubleshoot an aircraft hydraulics system. HYDRIVE supports generalized claims about aspects of student proficiency through probability-based combination of rule-based evaluations of specific actions. The paper highlights the interplay among inferential issues, the psychology of learning in the domain, and the instructional approach upon which the ITS is based.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Bayesian inference networks, cognitive diagnosis, HYDRIVE, intelligent tutoring systems, probability-based inference, student models.	41
	16. PRICE CODE N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

The Role of Probability-Based Inference in an Intelligent Tutoring System

Robert J. Mislevy & Drew H. Gitomer

Educational Testing Service

October, 1995

To appear in *User Modeling and User-Adapted Interaction*, special issue on numerical methods of handling uncertainty.

Copyright © 1995. Educational Testing Service. All rights reserved

The Role of Probability-Based Inference in an Intelligent Tutoring System

Robert J. Mislevy & Drew H. Gitomer

Educational Testing Service

Abstract

Probability-based inference in complex networks of interdependent variables is an active topic in statistical research, spurred by such diverse applications as forecasting, pedigree analysis, troubleshooting, and medical diagnosis. This paper concerns the role of Bayesian inference networks for updating student models in intelligent tutoring systems (ITSs). Basic concepts of the approach are briefly reviewed, but the emphasis is on the considerations that arise when one attempts to operationalize the abstract framework of probability-based reasoning in a practical ITS context. The discussion revolves around HYDRIVE, an ITS for learning to troubleshoot an aircraft hydraulics system. HYDRIVE supports generalized claims about aspects of student proficiency through probability-based combination of rule-based evaluations of specific actions. The paper highlights the interplay among inferential issues, the psychology of learning in the domain, and the instructional approach upon which the ITS is based.

Key words: Bayesian inference networks, cognitive diagnosis, HYDRIVE, intelligent tutoring systems, probability-based inference, student models

Overview

Intelligent tutoring systems (ITSs) depend on some form of student modeling to guide tutor behavior. Inferences about a student's current skills, knowledge, and strategy usage can affect the presentation and pacing of problems, the quality of feedback and instruction, and the determination of when a student has completed some set of tutorial objectives. But we cannot directly observe what a student does and does not *know*; this we must infer, imperfectly, from what a student does and does not *do*. This paper discusses an integration of principles of cognitive diagnosis and principles of probability-based inference in a framework for student modeling in intelligent tutoring systems.

Central to the development is the notion of the "student model", a set of variables corresponding to aspects of skill and knowledge that are important in the domain. Configurations of values of student-model variables approximate the multifarious skill and knowledge configurations of real students. There could be one or hundreds of variables in a student model. They could be categorical, qualitative, or numerical; they might concern tendencies in behavior, conceptions of phenomena, availability of strategies, or levels of aspects of developing expertise; they might be conceived as persisting over time or apt to change at the next problem step. The factors determining the form of the student model in a particular application are the nature and acquisition of competence in the domain, and the goals and philosophy of the instructional component of the system. The student model mediates between students' unique actions in specific situations, and the more abstract level of theory about the development of competence and the design of instruction.

Probability theory provides powerful mechanisms for explicating relationships, criticizing and improving models, and handling evidentiary subtleties, when it is possible to construct a joint distribution of variables whose modeled interrelationships approximate beliefs about the interrelated aspects of the real-world situation of interest—in this case, students' competencies and actions. Due to the recent developments sketched below, this requirement is not as constraining as is often believed. Discussions of the advantages of the probabilistic approach, compared to alternatives such as fuzzy logic and rule-based reasoning, appear in Cheeseman (1986), Pearl (1988), Schum (1979, 1994), and Spiegelhalter (1989). Two appealing features of probability-based reasoning for ITSs are its capabilities for principled synthesis of information from multiple, complex-structured observations, and for projecting beliefs about student-model variables to expectations for future observations, which can then be used for instructional decisions and, when

compared with actual observations, for model improvement. The viability of probability-based reasoning for expert systems in general sets the stage for investigating the scope and the limitations of the learning domains, student models, and instructional approaches for which probability-based reasoning can be profitably employed in the ITS context.

To this end, this paper discusses the implementation of probability-based reasoning in the HYDRIVE tutoring/assessment system for developing troubleshooting skills for the F-15 aircraft's hydraulics systems (Gitomer, Steinberg, & Mislevy, 1995). In the course of implementing principles of cognitive diagnosis, HYDRIVE uses a Bayesian inference network to express and update student-model variables—even as rule-based inference plays a complementary role in the system. Our objective is to share our experiences to date in exploring the ways that probability's conceptual and practical tools can be exploited in this context. We begin with an introduction to HYDRIVE that concentrates on its cognitive underpinnings, then review the basic elements of probability-based reasoning. Discussion of further developments in probability-based reasoning and the considerations they entail in HYDRIVE are interleaved in the presentation.

An Introduction to HYDRIVE

The hydraulics systems of the F-15 aircraft are involved in the operation of flight controls, landing gear, the canopy, the jet fuel starter, and aerial refueling. HYDRIVE simulates many important cognitive and contextual features of troubleshooting the F-15 hydraulics systems on the flightline. A problem starts with a video sequence in which a pilot, who is about to take off or has just landed, describes some aircraft malfunction to the hydraulics technician; for example, the rudders do not move during pre-flight checks. HYDRIVE's interface allows the student to perform troubleshooting procedures by accessing video images of aircraft components and acting on those components; to review on-line technical support materials, including hierarchically organized schematic diagrams; and to make instructional selections at any time during troubleshooting, in addition to or in place of instruction the system itself recommends. HYDRIVE's *system model* tracks the state of the aircraft system, including the fault to be isolated and any changes brought about by user actions. In a manner described below, the student's performance is monitored by evaluating how he or she uses available information about the system to direct troubleshooting actions. Components of HYDRIVE's *student model* diagnose the quality of specific troubleshooting actions, and characterize student understanding in terms of more general constructs such as knowledge of the systems, strategies, and procedures.

The rationale for HYDRIVE's design was established through the application of the PARI cognitive task analysis methodology (Means & Gott, 1988; Gitomer et al., 1992). These analyses were intended to reveal critical cognitive attributes that differentiate proficient from less-proficient performers in the domain of troubleshooting aircraft hydraulics systems. PARI tracing is a structured protocol analysis in which technicians are asked to solve a problem mentally, at each step detailing the reasons (**P**recursor) for the **A**ction that they would take. They are presented a hypothetical **R**esult and asked to **I**nterpret how the result modifies their understanding of the problem. They are also asked to represent their understanding of the specific aircraft system by drawing a block diagram of the suspect system. Differences appeared in three fundamental and interdependent areas, all of which seem necessary for an effective mental model for troubleshooting: system understanding, strategic understanding, and procedural understanding (Kieras, 1988).

System understanding. System understanding consists of how-it-works knowledge about the components of the system, knowledge of component inputs and outputs, and understanding of system topology, all at a level of detail necessary to accomplish necessary tasks. Novices' block diagrams did not evidence appropriate mental models of any hydraulics system sufficient to direct troubleshooting behavior. Experts evidenced a fuller understanding of how individual components operated within any given system (even though they did not understand the internal workings of these same components, which they had only to replace). Experts also demonstrated a principled sense of hydraulics system functioning beyond the specifics of the F-15, and organized their knowledge hierarchically according to the functional boundaries of the system. They understood the individual and shared characteristics of flight control and other hydraulics-related aircraft systems. An important consequence of this type of understanding is that, in the absence of a completely pre-specified mental model of a system, experts can construct a mental model using schematic diagrams. They can flesh out the particulars from their basic functional understanding of how hydraulics systems work in aircraft.

Strategic understanding. Novices did not employ effective troubleshooting strategies. That is, they demonstrated little ability to perform actions that would allow them to draw inferences about the problem from the behavior of the system. In many cases, the only strategy available to these individuals was to follow designated procedures in technical materials (Fault Isolation Guides, or FIs), even when it wasn't clear that the symptom matched the conditions described therein. While FIs can be useful tools, novices often fail to understand what information about the system a particular FI procedure provides or how

it serves to constrain the problem space. Even in those cases where they exhibit some system understanding, they frequently use a *serial elimination* strategy, wherein adjacent components are operated on in order. This strategy allows the technician to make claims only about a single component at a time. Experts try to use *space-splitting* strategies, isolating problems to a subsystem by using relatively few and inexpensive procedures that can rule out large sections of the problem area. When experts consult the FI guide, they do so as a reference to check whether they may be overlooking a particular problem source, and any FI action is immediately interpreted in terms of, and integrated with, their mental model of the system. Technicians with intermediate skills are quite variable in their use of strategies. When such individuals have fairly good system understanding for a specific situation, they frequently evidence effective troubleshooting strategies. When their system understanding is weak, they default to FI and serial elimination strategies.

Procedural understanding. Every component can be acted upon through a variety of procedures that provide information about some subset of the aircraft. Information about some types of components can only be gained by *removing and replacing* (R&R) them. Others can be acted upon by inspecting inputs and outputs (electrical, mechanical, and/or hydraulic), and by changing states (e.g., switches on or off, increasing mechanical input, charging an accumulator). Some actions, including most R&R procedures, provide information only about the component being acted upon, while other actions can provide information about larger pieces of the problem area under certain states of the system model. Novices are generally limited to R&R actions and the procedures specified in the FI. They often fail to spontaneously use the information that can be provided from studying gauges and indicators and conventional test equipment procedures. As individuals gain expertise, they develop a repertoire of procedures that can be applied during troubleshooting. Experts are particularly adept at partially disabling aircraft systems and isolating major portions of the problem area as functional or problematic.

The relationship between system, strategic, and procedural understanding. A mental model includes information not only about the inputs and outputs of components, but also about available actions that can be performed on components. The tendency to engage in certain procedures or strategies is often a function of the structure and completeness of system understanding, rather than the understanding of strategies or procedures in the abstract. A student's failure to execute a space-splitting action may appear at first to be a strategic failure, but the difficulty may lie with an impoverished understanding of the subsystem—a distinct possibility if the student has exhibited strong

strategic practice on other problems for which good system understanding exists. This view of troubleshooting expertise has implications for instruction as well as for inference. HYDRIVE's instruction focuses on effective system understanding and troubleshooting strategies rather than on optimizing actions to take at a given point in a problem. The instructional approach is to develop an understanding of the system as an hierarchy of interrelated models, a critical feature of expert knowledge, along with the general strategy of space-splitting in this system. HYDRIVE attempts to make this structure explicit through the use of hierarchical diagrams and similarly-organized verbal information.

Probability-Based Inference

When we reason from what we know and observe to explanations, conclusions, or predictions, the information we work with is typically incomplete, inconclusive, and amenable to more than one explanation (Schum, 1994). We attempt to establish the weight and coverage of evidence, as it informs the inferences and decisions we wish to make. While workers in every field address these questions as they arise with the kinds of inferences and the kinds of evidence they customarily address, interest in principles of inference at a level that might transcend the particulars of fields and problems has been keenest in the fields of statistics, philosophy, and jurisprudence. We focus on the concepts and the uses of mathematical or Pascalian probability-based reasoning, from what is usually called a subjectivist or personalist perspective (de Finetti, 1974; Savage, 1961).

A friend's request for advice on games of chance sparked Blaise Pascal's trailblazing application of the tools of mathematics to reasoning under uncertainty. He, followed by Bernoulli, Laplace, and others, laid out a framework for reasoning in such contexts. A "random variable" X is defined in terms of a collection of possible outcomes (the sample space), and a mapping from events (subsets of the sample space) to numbers which correspond to how likely they are to occur (probabilities). We will denote by $p(x)$ the mapping from a particular value x of X onto a probability. Probabilities satisfy the following requirements: (i) an event's probability is greater than or equal to 0, (ii) the probability of the event that includes all possible outcomes is 1, and (iii) the probability of an event defined as the union of two disjoint events is the sum of their individual probabilities. These simple axioms lead to consistent inference even for very complex situations, such as games with unknown probabilities linked in complicated ways or with events whose probabilities depend on the outcomes of earlier observations (a form of "conditional" probabilities, or the probability of x given that another variable Z takes the

value z, denoted $p(x|z)$—all of which can be verified empirically in repeatable chance situations such as games.

The applicability of mathematical probability for these aleatory, or chance, situations, is unquestioned. However, "there has been lingering controversy ... about the extent to which we should accept the Pascalian system ... as guides to life in probabilistic inference, especially when our evidence and hypotheses refer to singular or unique events whose probability can rest on no overt enumerative process" (Schum, 1994, p. 222). The personalistic Bayesian position is that if one's beliefs about a real-world situation are represented in the form of probability distributions, the axioms of mathematical probability ensure that all aspects of the individual beliefs are consistent with one another, or "coherent." This is particularly important when one must revise beliefs in response to new information—which is, after all, what student modeling in an ITS is all about. The real question is not whether probability-based reasoning is permissible in applications that lie outside the realm of repeatable chance situations, but the degree to which the salient aspects and relationships in a given real-world problem can be satisfactorily approximated in this framework. The following sections address issues encountered in defining variables, expressing their interrelationships, constructing suitable probability distributions, and carrying out inference, as they arise in the context of HYDRIVE.

Defining Variables In HYDRIVE

Unlike bridge hands and coin flips, few real-world problems present themselves to us in terms of natural "random variables." Random variables are not features of the world, but features of the patterns through which we organize our thinking about the world. From unique events, we must create abstractions which capture aspects we believe are salient but neglect infinitely many others. We must choose the level of detail at which variables will be defined, relationships will be modeled, and analyses will be carried out. Although probability and statistics textbooks *start* with predefined random variables, conceptualizing our problem in terms of variables amenable to probabilistic inference (particularly "observable variables") was one of the toughest challenges we faced!

Wenger (1987) describes three levels of information that student modeling might address in an ITS, and therefore at which variables can be defined. The *behavioral level* is often concerned with the correctness of student behaviors as compared with a model of expert performance (e.g., Brown, Burton & Bell's (1975) SOPHIE-I contrasted student behaviors with domain performance simulations in order to provide corrective feedback).

The *epistemic level* is concerned with particular knowledge states of individuals (Lesgold et al.'s (1992) SHERLOCK makes inferences about the goals and plans students are using to guide their actions during problem solving, and feedback is meant to respond to "what the student is thinking"; also see Appelt & Pollack, 1992, and Bauer, in press). The *individual level* addresses broader assertions about the individual that transcend particular problem states. Whereas the epistemic level of analysis might lead to the inference that "the student has a faulty plan for procedure X", the individual level of information might include the assertion that "the student is poor at planning in contexts that have properties A and B."

HYDRIVE aims to support generalized claims about aspects of student troubleshooting proficiency on the basis of detailed epistemic analysis of specific actions within the system. By bridging the gap between the individual and epistemic levels of information, the ITS is designed to have both the specificity to provide immediate feedback in a problem-solving situation, and the generality to help sequence problems, adapt instruction, and track proficiency in broad terms.

"Strategic knowledge", for example, is an abstraction that instructors use to summarize patterns of trainees' behavior—in conversations and classroom activities, as well as in their troubleshooting actions. We might therefore propose a variable called "strategic knowledge" for our student model, with possible values that represent increasing levels of expertise. Figure 1 depicts three possible states of belief about a student's "strategic knowledge." The first panel represents belief about a new student entering our course, reflecting our experience that most entering students are relatively weak in troubleshooting strategies. The second panel represents strong belief that a student is fairly good at troubleshooting strategies, a belief acquired perhaps from studying his transcript, reading his supervisor's recommendation, or observing expert-level troubleshooting actions. The third panel represents certainty that the student's level of expertise is "weak." Although a student's state of knowledge is never known with certainty, we shall see its role for reasoning in a "what if?" manner when structuring our knowledge about a domain. Later, we will pin down the meaning of "strategic knowledge" by specifying the tendencies of actions we expect in various troubleshooting situations from a student at each level, moderated by other student-model variables such as subsystem and procedural knowledge. These specifications represent deductive reasoning, from individual-level variables in the student model to probabilities of interpretations of observable actions (see de Rosis et al., 1992, and Jameson, 1992, for probability-based reasoning in similarly structured systems of person-level and observable variables).

As a student works through a HYDRIVE problem, the inferential task is to reason from the student's actions to implications in the student-model space. This problem is harder than the one faced in traditional educational assessment, since there predetermined observational settings with predetermined response categories (i.e., test items) can be devised and presented to students. Constraining observations in this manner limits what can be learned about students, but it is easy to know how to 'score' their responses. In a relatively unconstrained ITS such as HYDRIVE, however, students can take an unlimited number of routes through a problem. There are no clearly defined and replicable 'items' to score and calibrate. Different students carry out different sequences of action under different system-model configurations; each action depends on multiple aspects of competence, intertwined throughout the diverse situations students lead themselves through. We must, in some fashion, attempt to capture key aspects of their performance in terms consonant with the theory of performance that emerged from the cognitive analysis.

As an example, we may define a variable at a lower level of abstraction than "strategic knowledge": an "interpreted action" in a given problem situation. Interpreted actions lie at the epistemic level, taking the form of "plan recognition." Action sequences are not predetermined and uniquely defined in the manner of usual assessment items, since a student could follow a virtually infinite number of paths through the problem. Rather than attempting to model all possible system states and specific possible actions within them, HYDRIVE posits equivalence classes of states, or scenarios, each of which could arise many times or not at all as a given student works through a problem. The values of interpreted action variables are produced by HYDRIVE's system model, action evaluator, and strategy interpreter. The student activates the system model by providing input to the components; it processes the actions of the student and propagates sets of inputs and outputs throughout the system; the student can then examine the results for any other component of the system. The action evaluator calculates the action sequence's effects on the active problem area, so that a student's actions can be evaluated in terms of the information they yield in light of the previous actions.

For a given equivalence class of situations in which power-path splitting is possible, the potential values of interpreted action might be "power-path split", "serial elimination", "redundant action", "irrelevant action", and "remove and replace"—the value to be determined by the relationship of the effect of the action sequence on the problem area, as defined through information available to the student up through the time the action is taken. If, having supplied inputs, a student observes the output of a certain component

that the system model 'knows' is normal, then it is possible for the student to infer that all components on the activated path are functioning correctly and remove them from the problem area. If the student makes this interpretation and draws the appropriate inferences, then the problem areas that the student and HYDRIVE's system model hold will correspond and troubleshooting will continue with acceptable actions such as serial elimination and R&R, or expert actions, such as space-splitting, predominating. If the student incorrectly concludes that the observed component output was unexpected, then, in the student's mind, all the components in the active path remain in the problem area, others might be spuriously eliminated, and the problem area in the student's mind would begin to diverge from the one maintained by HYDRIVE; irrelevant and redundant actions become more likely.

The strategy interpreter employs a relatively small number of strategy interpretation rules (~25) to characterize the student's apparent strategy usage on the basis of the nature and the span of problem area reduction. An example of a student strategy rule is:

> *IF an active path which includes the failure has* ***not*** *been created and the student creates an active path which does* ***not*** *include the failure and the edges removed from the problem area are of one power class, THEN the student strategy is* ***splitting the power path.***

We note that these rules can be generalized to other troubleshooting domains. The generalizability resides in explicitly defining strategies in terms of an action's effect on the active problem area. While other domains may require different strategy definitions from HYDRIVE's, generalization is straightforward as long as these strategies can be referred to changes in the state of the problem area, or some similar representation.

Interpreted actions are examples of what are called "virtual evidence" in the expert systems literature; since students' plans are not actually observed, but are fallible judgments from the rule-based parsing of students' behaviors, there can be discrepancies between students' actual and interpreted reasons for actions. Plan recognition is most successful when both tasks and user actions are constrained, and plausible hypotheses about the space of potential plans are predetermined (e.g., Corbett & Anderson, 1995; Desmarais et al., 1993), because these factors reduce the uncertainty about students' plans given their actions. The uncertainty increases as constraints are relaxed, and as less can be anticipated about likely plans. At the limit, uncertainty in inferences about students' reasoning from single action sequences can render an ITS's feedback meaningless and its decisions misguided. For this reason, HYDRIVE's main instructional actions lie not at the level of

plan recognition, but at the level of accumulating patterns of interpreted actions. HYDRIVE uses a simpler rule-based logic to scan raw behavior for meaningful features without attempting the daunting and, for its purposes, pointless task of comprehensively explaining each one; it uses the more complex probability-based reasoning, as described below, to synthesize their meaning for more important instructional decisions.

Interrelationships Among Variables

While the terms "deductive", "inductive", and "abductive" inference have been used in somewhat different ways by different writers, Schum (1994) proposes definitions that are particularly useful for discussing the construction, utilization, and evolution of probability-based inference networks. *Deductive reasoning* flows from generals to particulars, within an established framework of relationships among variables—from causes to effects, from diseases to symptoms, from a student's knowledge and skills to observable behavior. *Inductive reasoning,* as Schum uses the term, flows in the opposite direction, also within an established framework of relationships—from effects to possible causes, from symptoms to probable diseases, from students' solutions or patterns of solutions to likely configurations of knowledge and skill; *abductive reasoning* proceeds from observations to new hypotheses, new variables, or new relationships among variables. Using this terminology, it may be said that Bayesian inference networks erect a reasoning structure in terms of deductive relationships, which, since the mathematical probability axioms are satisfied, supports coherent inductive inference. Model construction (developing a theory from which to posit variables and their interrelationships) and model improvement (modifying the network in response to unexpected or unsatisfactory outcomes) require abductive reasoning.

The theories and explanations of a field suggest the structure through which deductive reasoning flows. The requisite structure for deductive reasoning in HYDRIVE's student modeling emanates from the cognitive analyses: If a student is fairly familiar with troubleshooting strategies and the hydraulics system, but hazy about the workings of the landing gear system, what are the chances of various possible actions for a given state of a canopy failure? Inductive reasoning (in Schum's sense) flows through this same structure, but in the opposite direction: If a student makes a redundant action in a given state of a canopy failure, what does this imply about his familiarity with troubleshooting strategies, the hydraulics system, and the workings of the landing gear system? We will now render precise a simple exemplar relationship between a student-model variable and an interpreted-

action variable, and use it to illustrate probability-based deductive and inductive inference. Bayes' Theorem and the concepts of conditional dependence and independence are introduced in this connection. This will be followed by a discussion of how more complex interrelationships among many variables are represented in Bayesian inference networks.

Suppose that a student in question has strong knowledge of the problematic subsystem and relevant procedures, so that only strategic knowledge is at issue. In a situation near the problem solution, where space-splitting is no longer an option, what are our expectations that a student at each level of strategic knowledge might perform action sequences interpreted as "serial elimination", "redundant action", "irrelevant action", and "remove and replace"? Serial elimination is the best strategy available; remove and replace is useful but inefficient; both redundant and irrelevant actions are undesirable. Table 1 gives illustrative numerical values for probabilities of these actions at the different levels of proficiency. Figure 2 illustrates this flow of deductive reasoning. Each panel depicts the conditional probabilities of the various action categories, given level of strategic knowledge. We see increasing likelihood for serial elimination and decreasing likelihood of redundant and irrelevant actions as level of knowledge increases—although even experts sometimes make redundant moves, and novices sometimes make what appear to be expert moves, if not always for the same reasons experts make them.

Where do these probabilities come from? Initial values were set on the basis of qualitative input from expert instructors, patterns observed in PARI traces, and modifications based on "reasonableness checks" from simulated inputs and outputs (see von Winterfeldt & Edwards, 1986, on techniques of eliciting conditional probability distributions from subject matter experts). Current research in probability-based reasoning addresses modeling sources of information about these conditional probabilities, and the sensitivity of inferences to errors or misspecification in them. The probability framework also allows conditional probabilities to be characterized as unknown parameters—another level of modeling to represent our beliefs about the structures of relationships among observable and student-model variables—which can capture the "vagueness" of our beliefs about them, yet be coherently updated and made more precise as experience accumulates (Spiegelhalter et al., 1993). Whereas Table 1 simply provided numerical values for the conditional probabilities, a more complete representation of belief would take the form of a probability distribution for these conditional probabilities, which would itself depend on other aspects of knowledge and information.

In practice, we reason in the reverse direction: in ITSs, from interpreted actions to updated beliefs about students' strategic knowledge. This is accomplished in probability-based reasoning by means of Bayes' Theorem. Let X be a variable whose probability distribution $p(x|z)$ depends on the variable Z. Suppose also that prior to observing X, belief about the value of Z can be expressed in terms of a probability distribution $p(z)$ For example, we may consider all possible values of Z equally likely, or we may have an empirical distribution based on values observed in the past. Bayes' Theorem says

$$p(z|x) = p(x|z)p(z)/p(x), \quad (1)$$

where $p(x)$ is the expected value of $p(x|z)$ over all possible values of Z—a normalizing constant required by the axiom that belief about Z after having learned x must be represented by a probability distribution that sums to one. Suppose we start from the initial new-student beliefs about strategic knowledge from the first panel in Figure 1, and observe one action in the scenario that has the expectations depicted in Figure 2. The first panel in Figure 3 shows expectations for an action before it is observed; probabilities for the action variable are the average over the possible values of Strategic Knowledge, weighted by the initial belief probabilities for those possibilities. If we observe an action sequence interpreted as serial elimination and apply Bayes' Theorem, we obtain the results in the second panel of Figure 3. Because serial elimination is more likely to be carried out by students at higher levels of Strategic Knowledge, belief has shifted upwards from the first panel. Similar calculations would lead to the results in the remaining panels if we had observed any of the other possible interpretations.

This sequence illustrates the essence of the characterization of belief and of weight of evidence under the paradigm of mathematical probability (Good, 1950):

- Before observing a datum x, belief about possible values of a variable Z is expressed as a probability distribution, the prior distribution $p(z)$. The "prior" distribution can be conditional on other previous observations, and belief about Z may have been revised many times previously; the focus here is just on change in belief associated with observing x, *ceteris paribus*.

- After observing x, belief about possible values of Z is expressed in terms of another probability distribution, the posterior distribution $p(z|x)$.

- The evidential value of the observation x is conveyed by the likelihood function $p(x|z)$, the factor that revises the prior to the posterior for all possible values of Z. One can examine the *direction* by which beliefs associated with any given z change in response to observing x (is a particular value of z now more probable or less probable than before?) and the *extent* to which they change (by a little or by a lot?).

Bayesian Inference Networks

HYDRIVE moves from the space of unique observations to a space of random variables by interpreting action sequences in terms of equivalence classes. The challenge is to synthesize, in terms of belief about student-model variables, the import of many such actions—some in equivalent scenarios and others not, perhaps involving different subsystems and aspects of strategic understanding, each allowing for the possibility that the interpreter's evaluation does not match the student's thinking. Mathematical probability provides tools for combining evidence within a substantively determined structure—provided that the crucial elements of the situation can be satisfactorily mapped into the probability framework. The first requirement is to express the things we wish to talk about in terms of variables, as discussed above in the context of HYDRIVE. The second is to express the substantive, theoretical, or empirical relationships we perceive among them in terms of structural relationships among probability distributions.

Applying Bayes' Theorem in its textbook form (Eq. 1) quickly becomes unwieldy as the number of variables in a problem increases. Research on probability-based inference in complex networks of interdependent variables, or Bayes nets, has been spurred by applications in such diverse areas as forecasting, pedigree analysis, and medical diagnosis. Interest centers on obtaining the distributions of selected variables conditional on observed values of other variables, such as likely characteristics of offspring of selected animals given characteristics of their ancestors, probabilities of disease states given symptoms and test results, or, in the case of an ITS, values of student model variables given observed behaviors (see Mislevy, 1994a, 1995; Martin & VanLehn, 1993; Villano, 1992).

The notions of conditional independence and dependence are critical in this regard. Two random variables X and Y are *independent* if their joint probability distribution $p(x,y)$ is simply the product of their individual distributions, or $p(x,y) = p(x)p(y)$; equivalently, $p(x|y) = p(x)$ and $p(y|x) = p(y)$. Knowing the value of one provides no information about the value of the other. X is *dependent* on Z if belief about values of X varies with values of Z, as denoted by the conditional distribution $p(x|z)$. For example,

the troubleshooting action (X) we expect depends on a student's level of strategic knowledge (Z). This notion is important because the evidential value of an observation may depend in complex ways upon the other items of evidence (Schum, 1994, p. 208). Random variables X and Y are *conditionally independent* given Z if beliefs about X and Y are unrelated once the value of Z is known, even if they would have been related otherwise; that is, $p(x,y) \neq p(x)p(y)$ but $p(x,y|z) = p(x|z)p(y|z)$. The troubleshooting action we observe in one scenario certainly influences what we expect in the next, but we might posit it would not if we knew the values of all the relevant skill and knowledge variables.

Structuring a Bayes net begins with a *recursive representation* of the joint distribution of a set of random variables $x_1,\ldots x_n$, or

$$p(x_1,\ldots,x_n) = p(x_n|x_{n-1},\ldots,x_1)p(x_{n-1}|x_{n-2},\ldots,x_1)\cdots p(x_2|x_1)p(x_1) = \prod_{j=1}^{n} p(x_j|x_{j-1},\ldots,x_1), \quad (2)$$

where the term for j=1 is defined as simply $p(x_1)$. A recursive representation can be written for any ordering of the variables, but one that exploits conditional independence relationships is useful because variables drop out of the conditioning lists. For example, if X_3 is conditionally independent of X_2 given X_1, then $p(X_3|X_2,X_1)$ simplifies to $p(X_3|X_1)$. A graphical representation of (2), or a directed acyclic graph (DAG), depicts each variable as a node; each variable has an arrow drawn to it from any variables on which it is directly dependent (its "parents"). Conditional independence corresponds to omitting arrows ("edges") from the DAG, thus simplifying the topology of the network. In the example just given, the arrow from X_2 to X_3 can be omitted, leaving only the arrow from X_1 to X_3.

The conditional independence relationships suggested by substantive theory and discovered empirically determine the topology of the network of interrelationships in a system of variables. If it is favorable, the calculations required for probability-based reasoning can be carried out efficiently even in very large systems, by means of strictly local operations (implicit applications of Bayes' Theorem) on small subsets of interrelated variables ("cliques") and their intersections. Discussions of construction and computation in Bayesian inference networks are found in Lauritzen and Spiegelhalter (1988), Neapolitan (1990), and Pearl (1988).

A Simplified HYDRIVE Bayesian Inference Network

Figure 4 is a DAG expressing the dependence relationships in a simplified version of the inference network for the HYDRIVE student model. The direction of the arrows represents the deductive flow of reasoning used to construct probability distributions that incorporate the depicted dependence structure. A joint probability distribution for all these variables can be constructed by first assigning a probability distribution to each variable which has no parents (in this example, there is only one: "overall proficiency"); then for each successive variable, assigning a conditional probability distribution to its possible values for each possible combination of the values of its parents. The values expressed in these assignments incorporate such patterns as conjunctive or disjunctive relationships, incompatibilities, and interactions among diverse influences. The probabilities depicted in Figure 4 correspond to the initial status of belief about all variables in the network, or before any actions are observed from a student. They are determined by the initial distribution for "overall proficiency" and the posited conditional probabilities for all other variables in the network given their parents.

Four groups of variables can be distinguished in Figure 4: (1) The rightmost nodes are the "interpreted actions", the results of rule-driven epistemic analyses of students' actions in a given situation. Two prototypical sets appear, each corresponding to an equivalence class of potential observables in a given scenario: canopy situations in which space-splitting is not possible, and landing-gear situations in which space-splitting is possible. Three members are represented from each class. (A virtual storage algorithm allows the full network to absorb information from an indefinite number of variables in such a class while storing and manipulating only two copies of representative class members; see Mislevy, 1994b.) (2) The immediate parents of the interpreted action variables are the knowledge and strategy requirements that in each case define the class. The possible values are all combinations of the values of the system and strategic knowledge variables that play a role in the scenario class, as indicated by the directed arrows into these nodes. There are too many to depict, so the node is left blank rather than showing all the probability bars. (3) The long column of variables in the middle concerns aspects of subsystem and strategic knowledge, which correspond to instructional options. We see that canopy actions in which space-splitting is not possible are conditionally independent of space-splitting proficiency, given the proficiencies that are directly relevant. (4) To the left are summary characterizations of more generally construed proficiencies.

Serial elimination is the best strategy in a canopy/no-split situation, and, as expressed in conditional probabilities that embody deductive reasoning in the network, is likely when the student has strong knowledge of this strategy and all relevant subsystems. Remove-and-replace is most likely when a student possesses some subsystem knowledge but lacks familiarity with serial elimination, whereas weak subsystem knowledge increases chances of irrelevant and redundant actions. Figure 5 depicts belief after observing, in three separate situations in the canopy/no-split class, one redundant and one irrelevant action (both ineffectual troubleshooting moves) and one remove-and-replace (serviceable but inefficient).

Subsystem and *strategy variables* serve to summarize tendencies in interpreted behaviors at a level addressed by instruction, and to disambiguate patterns of actions in light of the fact that inexpert actions can have several causes. Figure 5, which is posterior to three inexpert canopy actions, shows belief shifted from values in Figure 4, toward lower values for serial elimination and all subsystem variables directly involved in the situation (mechanical, hydraulic, and canopy knowledge). Any or all could be the source of the student's difficulty, since all are required for high likelihoods for expert actions. Belief about the student's level knowledge of subsystems not directly involved in these situations is also lower, because students unfamiliar with one subsystem tend to be unfamiliar with others; also, to a lesser extent, students unfamiliar with subsystems tend to be unfamiliar with troubleshooting strategies. These relationships are expressed through the more general *system* and *strategic knowledge* variables at the left of the figure. These variables serve to exploit the indirect information about aspects of knowledge not directly tapped in a given scenario, and to summarize broadly construed aspects of proficiency for purposes of evaluation and problem selection.

Figures 6 and 7 represent the state of belief that would result after further observing two different sets of actions in situations involving the landing gear in which space-splitting is possible. Figure 6 shows the results of three more inexpert action sequences. Status on all subsystem and strategy variables is further downgraded, and reflected in the more generalized summary variables. Figure 7 shows the results that would obtain if, instead, one observed three good actions: two space-splits and one serial elimination. Belief about strategic skill has shifted toward higher levels, as have beliefs about subsystems involved in the landing gear situations. Weakness in mechanical, hydraulic, and/or canopy subsystem knowledge are now the most plausible explanations of the three inexpert canopy situation actions. The diffuse belief at the generalized proficiency level results from the

uneven profile of subsystem knowledge. In this network, diffuse belief at higher levels in the student model can result either from lack of information about finer-grained aspects of the student's knowledge, or, as in this situation, from fairly accurate but conflicting information about them.

We did not have the luxury of large numbers of solutions from acknowledged experts and novices of various configurations, from which to determine the conditional probabilities of observable variables given student-model variables. Initial values were set on the basis of expert opinion and checked by means of data obtained in PARI traces. We have recently acquired traces of forty students working through ten problems each, from which we may empirically improve the original conditional probability specifications in the manner described by Spiegelhalter and Cowell (1992). With larger amounts of empirical data, we can capitalize on the probability framework to carry out formal statistical model-checking procedures. After two-thirds of a student's actions have been entered, for example, updated student-model parameters and conditional distributions yield predictive distributions for subsequent actions. These model-based predictive distributions can be compared with the actual remaining third of the observations to verify model calibration, or to provide clues for improving the model.

Additional Grounds for Revising Belief

In the preceding discussion and examples, observations obtained sequentially over time are presumed to simply provide additional information about unchanging values of student-model variables. The whole point of an ITS, however, is to help students change over time; in particular, to improve their proficiencies. This section concerns two additional reasons for modifying belief about student-model variables: change due to explicit instruction, and change due to implicit learning. In both cases, the requirement under a probabilistic approach is to do so in a manner that maintains coherence. We discuss an approach to accomplishing this end while avoiding the construction and maintenance of a full dynamic model.

Updating based on direct instruction. While HYDRIVE's system model functions as a discovery world for system and procedural understanding from the student's point of view, the evaluations its student-modeling components make are based on an implicit strategic goal structure observed in expert troubleshooting. This structure is made explicit in HYDRIVE's instruction. The student is given great latitude in pursuing the problem solution, with prompts or reminders given only when an action violates important rules

associated with the strategic goal structure. HYDRIVE recommends direct instruction only when information that accumulates across scenarios shifts belief about, say, knowledge of a subsystem or strategy, sufficiently downward to merit more specifically focused feedback, review, and exercises. In light of the compatibility of probability-based inference and decision theory, a natural extension of the system we have not yet undertaken would be to incorporate decision-theoretic reasoning to manage these interventions.

We expect direct instruction to change students' understanding. The change in our belief about the values of a student-model variable due to instruction differs from the previously discussed updating of belief about presumably static student-model variables due to observed actions. The change due to instruction might be modeled as dependent on the student's previous level of understanding and the expected effects of instruction, perhaps additionally informed by a posttest following the instruction. A fully specified dynamic model is schematized in the first panel of Figure 8, in which multiple time points, with corresponding multiple copies of student-model variables, are jointly modeled and maintained. Multiple copies of observable variables are also shown, with expectations that correspond to belief about possibly different values of student-model variables. As proficiency increases with instruction, for example, expectations for expert actions in classes of relevant situations increase.

A more parsimonious alternative to jointly modeling all variables before and after instruction employs a small stand-alone Bayesian network to account for change due to instruction. A single time-point network for the full set of student-model and observable variables is maintained, but variables affected by direct instruction are modified in accordance with this stand-alone network, replaced in the appropriate nodes, and implications propagated in the same manner as are changes effected by observations. The result is the "virtual" dynamic network schematized in the second panel of Figure 8. Figure 9 is an example of the stand-alone network. Table 2 gives the corresponding conditional probabilities; these can be refined over time, starting with expert opinion and limited experience but honed as experience accumulates. Conditional independence with respect to other student-model and observable variables is implied by the use of the stand-alone network. The probability distribution for the relevant student-model variable before instruction and the outcome of an instructional posttest exercise are entered, and the distribution posterior to instruction is obtained. The resulting posterior distribution for the student-model variable is replaced into the full network in a manner that assures coherence will be maintained,[1] and the consequences of this change are propagated through the

network in the usual manner in order to revise accordingly beliefs about other student-model variables and expectations about future observations.

Updating based on learning while problem-solving. Even without direct instruction, students can be expected to improve their troubleshooting skills as a result of practicing them and thinking through the problems. Although this probably occurs incrementally throughout a problem, we follow Kimball's (1982) expedient of revising belief due to implicit learning only at problem boundaries. Kimball's tutor, like Anderson's LISP tutor (Corbett & Anderson, 1995), revises belief in a manner consistent with probability axioms through an explicit learning model. That is, a particular functional form for change is presumed, and degree of learning is also assumed or estimated. We employ a more conservative and less model-bound approach: The ITS accommodates the student's learning by gradually discounting information from past actions that were determined by earlier, presumably lower, levels of understanding. The student learns; to account for this, the system that models his knowledge forgets.

The idea is to enter each problem with student-model variable distributions that generally agree with the final values from the previous problem as to direction and central tendency, but are more diffuse and thus easier to change in light of new actions driven by possibly different (presumably improved) values. Two strategies for accomplishing this end are (1) downweighting the influence of actions as they recede in time, and (2) between problem sessions, mixing then-current posterior distributions with noninformative distributions and propagating the revised versions through the network as described above for instructional revisions. These "decaying-information estimators" are less efficient than full-information estimators if there is no change over time, or if there is change and it is modeled accurately; but, when trends do exist, they can provide better approximations than either ignoring it or modeling it incorrectly.

Discussion

Mathematical probability provides powerful machinery for coherent reasoning about complex and subtle interrelationships—to the extent that one can capture within its framework the key aspects of a real-world situation. If this can be accomplished, advantages both conceptual and practical accrue. A Bayes net built around the generating principles of the domain makes interrelationships explicit and public, so one can not only monitor what one believes, but communicate why one believes it. A model can be refined over time in light of new information, as when initial subjective conditional probability

specifications are updated in light of accumulating data. Able to calculate predictive distributions of any subset of variables given values of any others, one can investigate a modeled structure by entering hypothetical data to check for fidelity to what one believes, or entering real data to check for fidelity to what one observes (see the review by Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993, on model-checking tools for complex networks). It may be painstaking and difficult work to carry out the requisite modeling tasks, but recent progress in calculation, model-building, and model-checking has been explosive (again, see Spiegelhalter et al., 1993).

The challenge most significant in any application of probability-based reasoning is channeling one's scope of vision from an open-ended universe of human experience, to a closed universe of variables and probability distributions. We experienced this constraint in HYDRIVE first in having to interpret observations in terms of variables over which probabilities sum to one. Just how to do this was not immediately obvious in HYDRIVE's unconstrained observational setting. We eventually cast interpreted actions as members of exhaustive and mutually exclusive classes, so that the updating that occurs when a space-split *did* occur depends intimately on the fact that an R&R (remove & replace), serial elimination, or redundant or irrelevant action *could have*, but *did not* occur. HYDRIVE's progenitor, SHERLOCK (Lesgold et al., 1992) also interprets action sequences in terms of inferred plans, but it changes values of student-model variables according to action-specific rules that address only inference from evaluated actions to student-model variables. These rules are easier to construct than HYDRIVE's conditional probability structures, because the rules triggered by any observation can be specified without regard to rules for other potential observations. But since no provision is made for reasoning from student-model values to future actions, claims of student proficiency are difficult to check conceptually or empirically. An interpreted action in SHERLOCK may be an "event" in the everyday sense of the word, but it is not in the sense of mathematical probability.

The constraints of mathematical probability also pinch in the presumption that all potential states of the real-world situation can be satisfactorily approximated under the model, relative to the purpose at hand. Shafer (1976) calls modeling the possibilities one will explore "defining the frame of discernment." But what if a particular student's conception differs from any of the postulated models? The probabilities that result from the use of Bayes' Theorem depend on the posited structure. Only possibilities built into the model can end up with positive probabilities! Apparently precise numerical statements of

belief prove misleading or downright embarrassing when it is later determined that the true state of affairs could not even be approximated in the analytic model.

Two strategies help address this problem in applied settings. One approach is to augment theoretically-expected unobservable states with one or more "catch-all" states which increase in probability when unexpected patterns arise in observable data (e.g., the class associated with flat likelihood for all symptom patterns in the MUNIN expert system for neuromuscular diseases, described in Andreassen et al., 1987; its posterior probability increases when symptoms appear that fail to match any of the patterns typical of the diseases built into the model). Another approach is to calculate indices of model misfit or "surprise" (e.g., Good, 1971). While carrying out inference within a given probabilistic structure, one calculates indices of how usual or unusual the observed data are under that structure. Both of these approaches can flag patterns of evidence that are not likely under *any* of the possibilities built into the model, calling for model revision (further abductive reasoning, in Schum's sense).

Conclusion

Probability-based reasoning has emerged as a viable approach to structuring and managing knowledge in the presence of uncertainty, due partly to computational advances such as rapid local updating (Spiegelhalter et al., 1993), but more to conceptual progress—a confluence of ideas about personal probability (e.g., Savage, 1961; de Finetti, 1974) and the structuring of inference (e.g., Schum, 1994). This progress was spurred by the emergence of alternative frameworks for reasoning in the presence of uncertainty, such as fuzzy logic (Zadeh, 1965) and the Dempster-Shafer theory of evidence (Shafer, 1976). Whether mathematical probability *couldn't* be used to deal with the problems that promoters of alternative approaches advanced was fiercely contested, but clearly it *wasn't*. We can safely predict continued rapid progress along statistical lines, increasing prospects for the usefulness of probability-based reasoning in intelligent tutoring systems.

Perhaps the main lesson we take from HYDRIVE is the importance of cognitive grounding. Arguing in the abstract about advantages and disadvantages of approaches to managing uncertainty is well and good, and quite necessary—but in the final analysis, the success of a given application will depend on identifying the key concepts and interrelationships in the domain. Ad hoc reasoning with sound substance beats coherent reasoning with inadequate substance, if you must choose between them—but coherent reasoning around sound substance dominates! Especially germane to the ITS context are

(1) understanding principles of the target domain and how people learn those principles, so as to structure the student model efficaciously, and (2) determining what one needs to observe, and how it depends on students' possible understandings, so as to structure observable variables and their relationship to student-model variables.

Concepts from statistics, cognitive psychology, and instructional science must come together for a successful ITS. Over time, prototypical approaches for developing ITSs consonant with the principles of these domains must evolve, in the form of examples, effective approaches to common problems, knowledge elicitation schemes aligned to the anticipated model, and expedients that strike good balances among competing properties such as fidelity and computability. Our experiences with HYDRIVE persuade us that the quest will be arduous, but worthwhile.

Notes

[1] For a single affected student-model variable X, this revision is accomplished as follows: Suppose that belief about X before instruction is expressed by probabilities $(b_1, b_2, \ldots, b_m)$ for its m possible values. These are initializing values in the stand-alone network. Instruction is provided and the posttest is administered; in accordance with the conditional probabilities in the stand-alone network, a revised vector of beliefs about X is obtained, say $(c_1, c_2, \ldots, c_m)$. The columns in the potential table in the full network into which evidence about X is absorbed are reweighted by the factors $(c_1/b_1, c_2/b_2, \ldots, c_m/b_m)$, so the resulting beliefs about X take the desired values $(c_1, c_2, \ldots, c_m)$. The consequences of entering this so-called "virtual evidence" are propagated throughout the rest of the network. This scheme can be extended to cases in which instruction directly affects multiple student-model variables. Coherent revision of joint beliefs is accomplished through the use of a new variable defined as the joint product of all pertinent individual student-model variables. This extended variable serves as the interface between the full and stand-alone nets in the manner described above for a single variable.

Acknowledgments

The examples in this paper are based on the HYDRIVE Intelligent Tutoring System, which has been supported by Armstrong Laboratories of the United States Air Force. We are indebted to Sherrie Gott and Ellen Hall and their staffs for their contribution to this effort. Mr. Mislevy's work was supported in part by Contract No. N00014-91-J-4101, R&T 4421573-01, from the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research, and by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Educational Research and Development Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The views expressed here are those of the authors and do not imply any official endorsement by any organizations funding this work. We are grateful to UMUAI Editor Alfred Kobsa, Guest Editor Anthony Jameson, three anonymous referees, Irv Katz, and Ming-Mei Wang, for their many thoughtful comments on earlier drafts. The Bayes net calculations reported here were carried out with the ERGO computer program (Noetic Systems, 1991).

References

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S.K. (1987). MUNIN: A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 366-372). San Mateo, CA: Morgan Kaufmann.

Appelt, D.E., & Pollack, M. E. (1992). Weighted abduction for plan ascription. *User Modeling and User-Adapted Interaction*, **2**, 1-25.

Bauer, M. (in press). A Dempster-Shafer approach to modeling agent preferences for plan recognition. *User Modeling and User-Adapted Interaction*, in this special issue.

Brown, J. S., Burton, R. R. & Bell, A. G. (1974). SOPHIE: A sophisticated instructional environment for teaching electronic troubleshooting. *BBN REPORT 2790.* Cambridge, MA: Bolt Beranek and Newman, Inc.

Cheeseman, P. (1986). Probabilistic versus fuzzy reasoning. In L.N. Kanal & J.F. Lemmer (Eds.), *Uncertainty in artificial intelligence* (pp. 85-102). Amsterdam: North-Holland.

Corbett, A.T., & Anderson, J.R. (1995). Model tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, **4**, 253-278.

de Finetti, B. (1974). *Theory of probability* (Volume 1). London: Wiley.

de Rosis, F., Pizzutilo, Russo, A., Berry, D.C., & Molina, F.J.N. (1992). Modeling the user knowledge by belief networks. *User Modeling and User-Adapted Interaction*, **2**, 367-388.

Desmarais, M.C., Giroux, L., & Larochelle, S. (1993). An advice-giving interface based on plan-recognition and user-knowledge assessment. *International Journal of Man-Machine Studies*, **39**, 901-924.

Gitomer, D.H., Cohen, W., Freire, L., Kaplan, R., Steinberg, L., & Trenholm, H. (1992). *The software generalizability of HYDRIVE* (Armstrong Laboratories Progress Report). Princeton, NJ: Educational Testing Service.

Gitomer, D.H., Steinberg, L.S., & Mislevy, R.J. (1995). Diagnostic assessment of trouble-shooting skill in an intelligent tutoring system. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 73-101). Hillsdale, NJ: Erlbaum.

Good, I.J. (1950). *Probability and the weighting of evidence.* London: Griffin; New York: Hafner.

Good, I.J. (1971). The probabilistic explication of information, evidence, surprise, causality, explanation, and utility. In V.P. Godambe & D.A. Sprott (Eds.),

Foundations of statistical inference (pp. 108-141). Toronto: Holt, Rinehart, and Winston.

Jameson, A. (1992). Generalizing the double-stereotype approach: A psychological perspective. *Proceedings of the Third International Workshop on User Modeling* (69-83), Dagstuhl, Germany, August 1992.

Kieras, D.E. (1988). What mental model should be taught: Choosing instructional content for complex engineered systems. In M.J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned* (pp. 85-111). Hillsdale, NJ: Lawrence Erlbaum.

Kimball, R. (1982). A self-improving tutor for symbolic integration. In D. Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems* (pp. 283-307). London: Academic Press.

Lauritzen, S.L., & Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B,* **50,** 157-224.

Lesgold, A. M., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. W. Regian and V.J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum.

Martin, J.D., & VanLehn, K. (1995). A Bayesian approach to cognitive assessment. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 141-165). Hillsdale, NJ: Erlbaum.

Means, B., & Gott, S.P. (1988). Cognitive task analysis as a basis for tutor development: Articulating abstract knowledge representations. In M.J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned* (pp. 35-58). Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1994a). Evidence and inference in educational assessment. *Psychometrika,* **59,** 439-483.

Mislevy, R.J. (1994b). Virtual representation of IID observations in Bayesian belief networks. *ETS Research Memorandum 94-13-ONR*. Princeton, NJ: Educational Testing Service.

Mislevy, R.J. (1995). Probability-based inference in cognitive diagnosis. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 43-71). Hillsdale, NJ: Erlbaum.

Neapolitan, R.E. (1990). *Probabilistic reasoning in expert systems: Theory and algorithms.* New York: Wiley.

Noetic Systems, Inc. (1991). ERGO [computer program]. Baltimore, MD: Author.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Kaufmann.

Savage, L.J. (1961). The foundations of statistics reconsidered. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics and Probability,* Vol. I (pp. 575-586). Berkeley: University of California Press.

Schum, D.A. (1979). A review of a case against Blaise Pascal and his heirs. *Michigan Law Review,* **77,** 446-483.

Schum, D.A. (1994). *The evidential foundations of probabilistic reasoning.* New York: Wiley.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton: Princeton University Press.

Spiegelhalter, D.J. (1989). A unified approach to imprecision and sensitivity of beliefs in expert systems. In L.N. Kanal, J. Lemmer, & T.S. Levitt (Eds.), *Artificial intelligence and statistics* (pp. 47-68). Amsterdam: North-Holland.

Spiegelhalter, D.J., & Cowell, R.G. (1992). Learning in probabilistic expert systems. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (Eds.), *Bayesian Statistics 4* (pp. 447-465). Oxford, U.K.: Oxford University Press.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., & Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science,* **8,** 219-283.

Villano, M. (1992). Probabilistic student models: Bayesian belief networks and knowledge space theory. In C. Frasson, G. Gauthier, & G. McCalla (Eds.), *Intelligent Tutoring Systems: Proceedings of the Second International Conference, ITS '92* (491-498), Montreal, June 1992.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge: Cambridge University Press.

Wenger, E. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge.* Los Altos, CA: Morgan Kaufmann.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control,* **8,** 338-353.

Table I

Numerical Values of Conditional Probabilities of Interpreted Action Sequences, Given Strategic Knowledge

	Conditional Probability of Interpreted Action Sequence			
Strategic Knowledge	Serial Elimination	Redundant Action	Irrelevant Action	Remove and Replace
Expert	.75	.10	.05	.10
Good	.50	.10	.10	.30
Okay	.30	.15	.15	.40
Weak	.20	.20	.30	.30

Table II

Conditional Probability Tables Concerning Strategic Knowledge after Instruction

Status before Instruction	Conditional Probability of Status after Instruction			
	Expert	Good	Okay	Weak
Expert	.95	.05	.00	.00
Good	.30	.65	.05	.00
Okay	.10	.40	.45	.05
Weak	.05	.30	.45	.20

Status after Instruction	Conditional Probability of Posttest Performance after Instruction		
	High	Medium	Low
Expert	.90	.10	.00
Good	.75	.25	.00
Okay	.25	.50	.25
Weak	.00	.25	.75

List of Figures

1. Three Configurations Representing Possible Belief About "Strategic Knowledge"

2. Conditional Probabilities of Interpreted Action Sequences, Given Strategic Knowledge

3. Updated Probablilities for Strategic Knowledge, Given Interpreted Action Sequences

4. Initial Status of Student Model (i.e., Before Observing Any Actions)

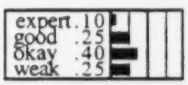
5. Status of Student Model After Observing Three Inexpert Actions in Canopy Situations

6. Status of Student Model After Observing Three Inexpert Actions in Canopy Situations.and Three Inexpert Actions in Landing Gear Situations

7. Status of Student Model After Observing Three Inexpert Actions in Canopy Situations and Three Expert Actions in Landing Gear Situations

8. Schematic Diagrams for Two Approaches to Dynamic Modeling

9. A Stand-Alone Network for Updating Belief about Strategic Knowledge due to Direct Instruction



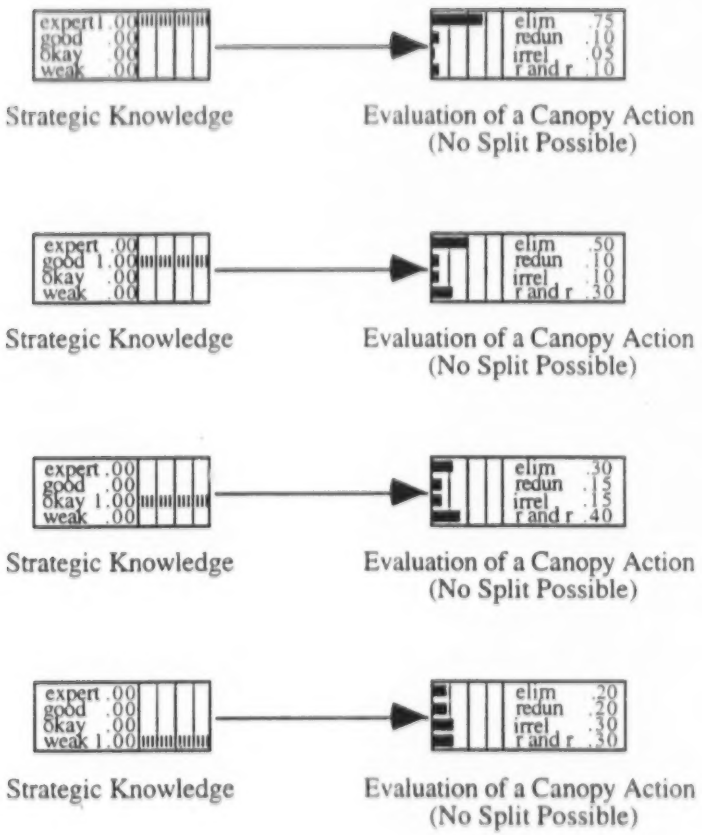
Strategic Knowledge

Strategic Knowledge

Strategic Knowledge

Figure 1

Three configurations representing possible belief about "Strategic Knowledge"



Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending the full width of a node represents certainty, due to having observed the value of that variable; i.e., a student's actual responses to tasks.

Figure 2

Conditional probablilities of interpreted action sequences, given Strategic Knowledge

a) Initial Status

Strategic Knowledge — Evaluation of a Canopy Action (No Split Possible)

b) Status after observing an action evaluated as serial elimination

Strategic Knowledge — Evaluation of a Canopy Action (No Split Possible)

c) Status after observing an action evaluated as redundant

Strategic Knowledge — Evaluation of a Canopy Action (No Split Possible)

d) Status after observing an action evaluated as irrelevant

Strategic Knowledge — Evaluation of a Canopy Action (No Split Possible)

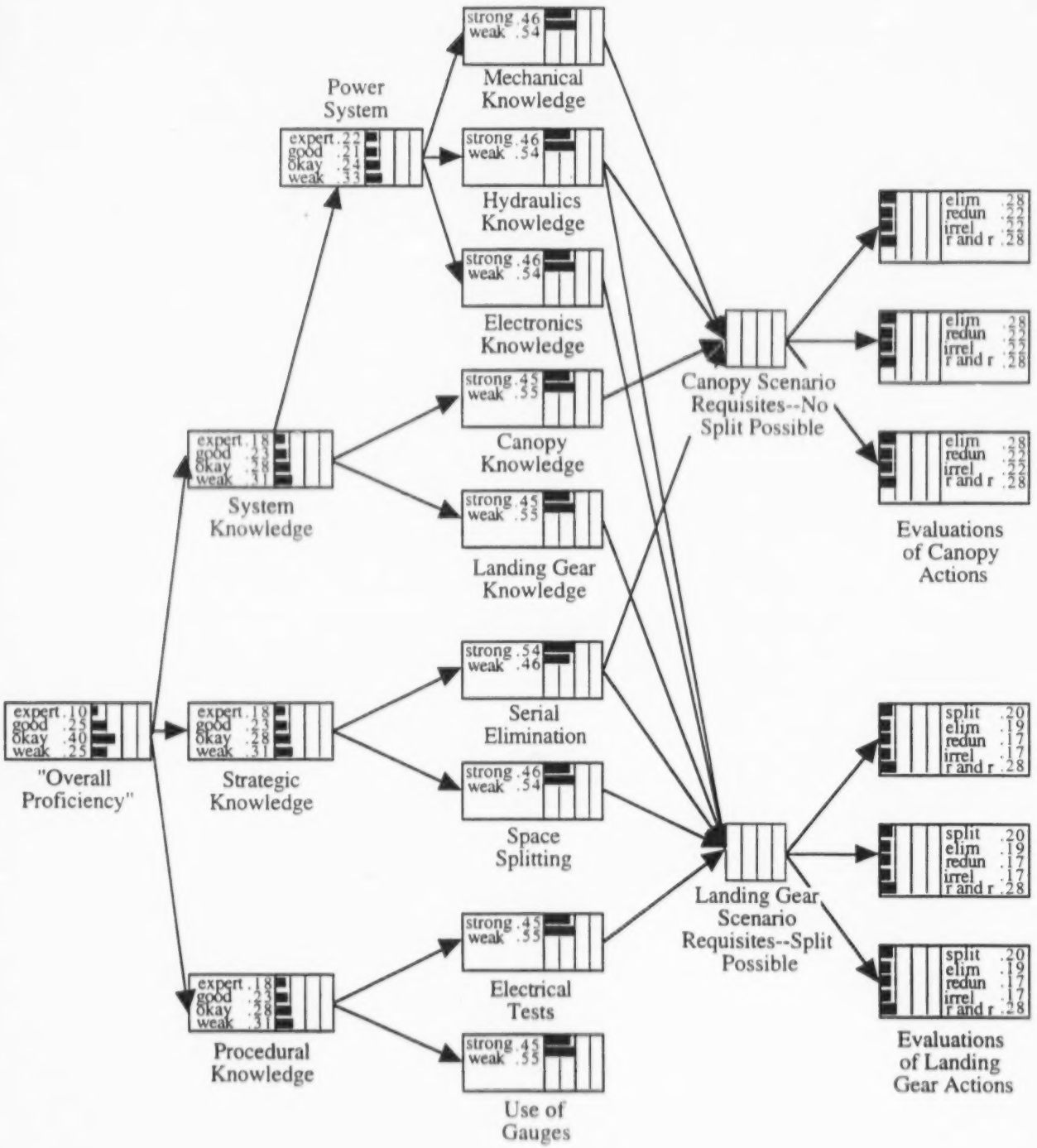
e) Status after observing an action evaluated as remove and replace

Strategic Knowledge — Evaluation of a Canopy Action (No Split Possible)

Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending the full width of a node represents certainty, due to having observed the value of that variable; i.e., a student's actual responses to tasks.

Figure 3

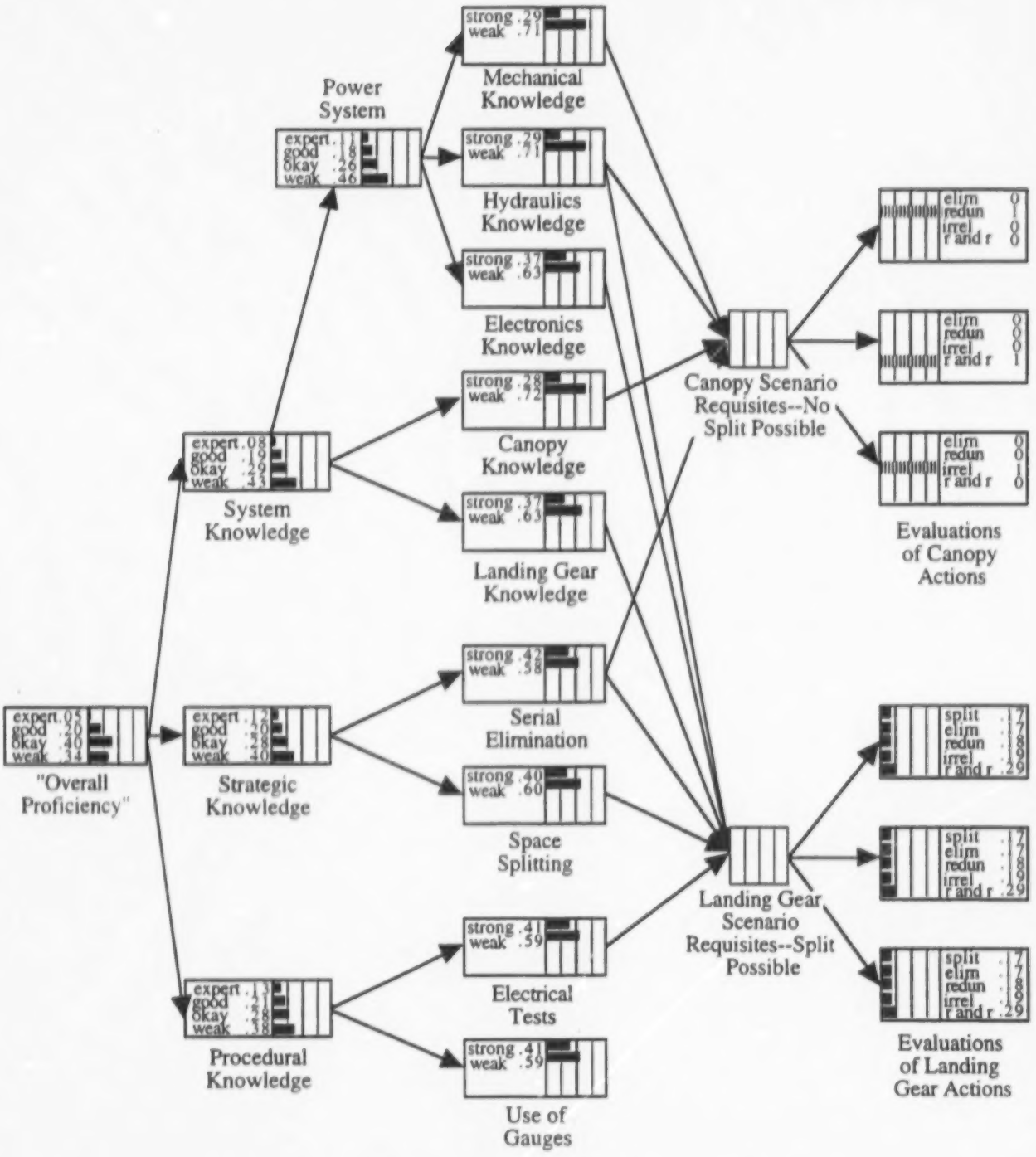
Updated probablilities for Strategic Knowledge, given interpreted action sequences



Note: Bars represent probabilities, summing to one for all the possible values of a variable.

Figure 4

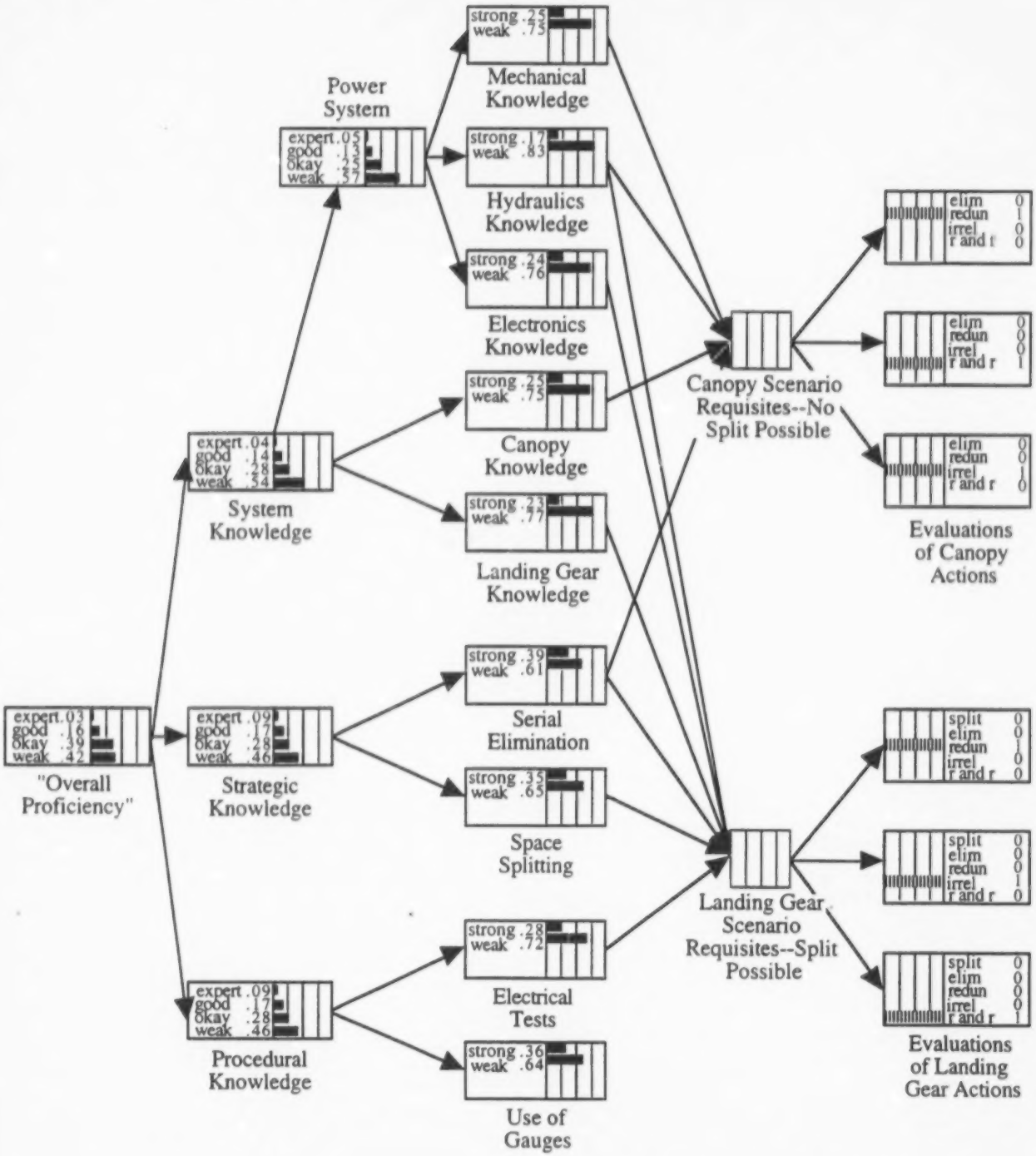
Initial Status of Student Model (i.e., Before Observing any Actions)



Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending the full width of a node represents certainty, due to having observed the value of that variable; i.e., a student's actual responses to tasks.

Figure 5

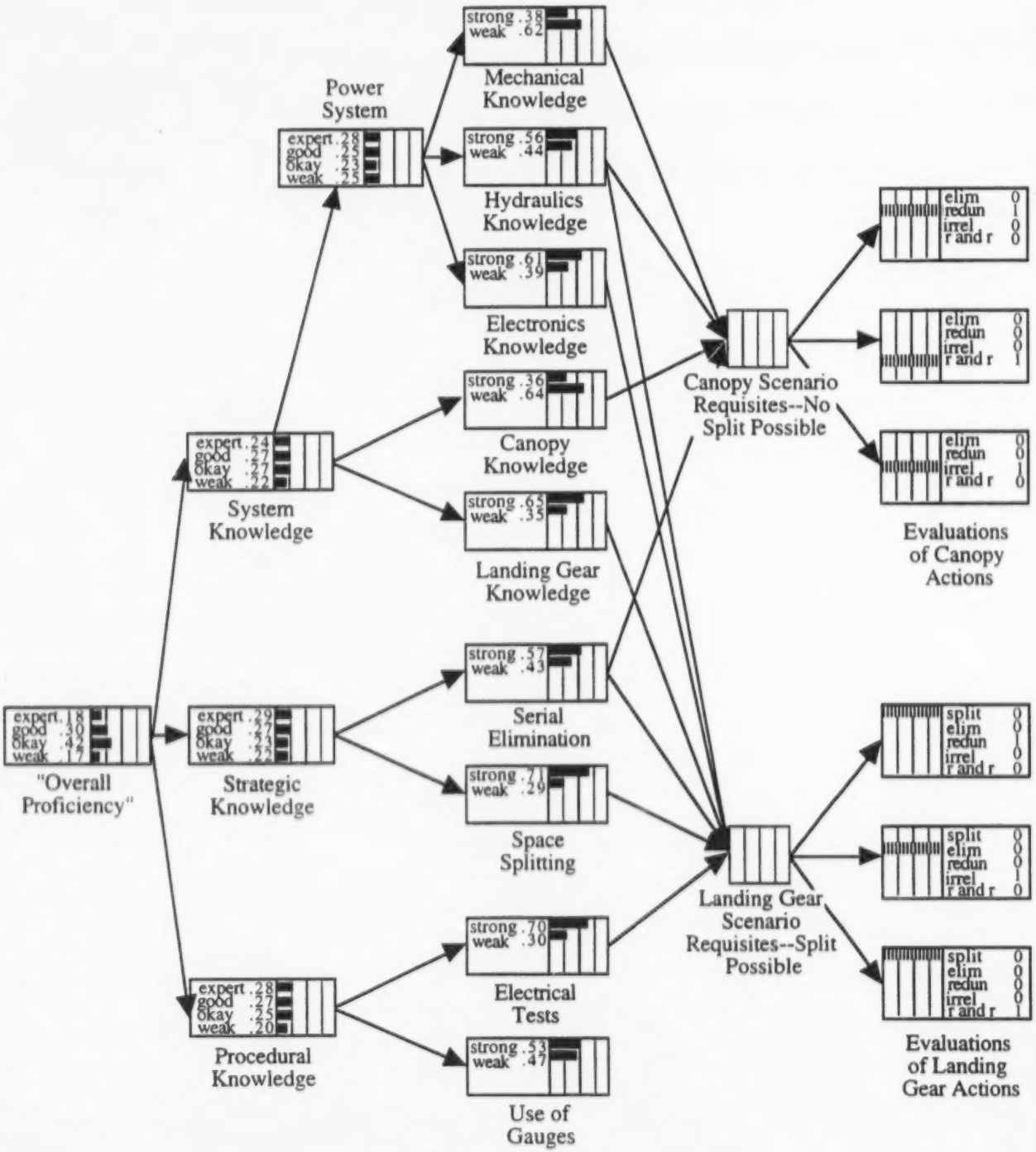
Status of Student Model after Observing Three Inexpert Actions in Canopy Situations



Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending the full width of a node represents certainty, due to having observed the value of that variable; i.e., a student's actual responses to tasks.

Figure 6

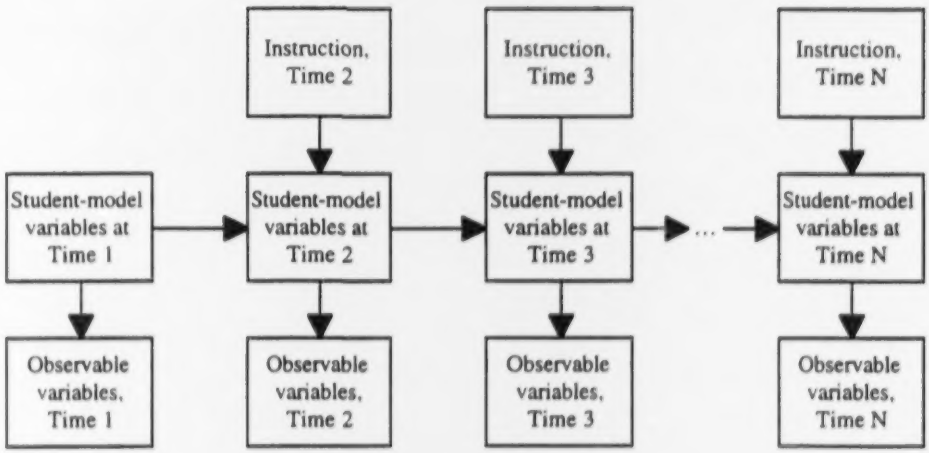
Status of Student Model after Observing Three Inexpert Actions in Canopy Situations and Three Inexpert Actions in Landing Gear Situations



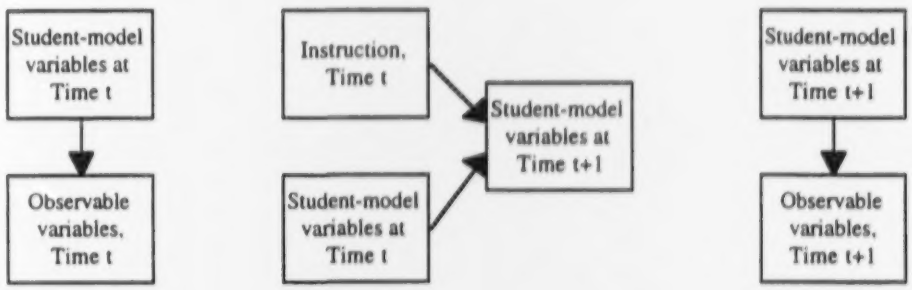
Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending the full width of a node represents certainty, due to having observed the value of that variable; i.e., a student's actual responses to tasks.

Figure 7

Status of Student Model after Observing Three Inexpert Actions in Canopy Situations and Three Expert Actions in Landing Gear Situations



A schematic outline for incorporating change to student-model variables, explicitly modeling all time points at once.


Student-model variables at Time t
Observable variables, Time t
Instruction, Time t
Student-model variables at Time t
Student-model variables at Time t+1
Student-model variables at Time t+1
Observable variables, Time t+1

A schematic outline for incorporating change to student-model variables, explicitly modeling only one time point at a time and updating student-model variables in a separate network

Figure 8

Schematic Diagrams for Two Approaches to Dynamic Modeling

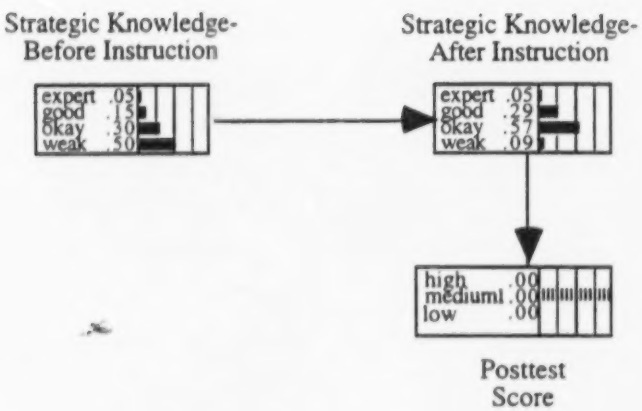


Figure 9

A Stand-Alone Network for Updating Belief about
Strategic Knowledge due to Direct Instruction

Distribution List

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Robert L Albright
Educational Testing Service
16-C Rosedale Road
Princeton, NJ 08541

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
School of Education
Tel Aviv University
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Dr Nancy S Cole
Educational Testing Service
14-C Rosedale Road
Princeton NJ 08541

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education Research
Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr. George Macready
Professor
Dept of Meas., Stat. & Eval.
College of Education
Room 1230C, Benjamin Bldg.
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Tech Training Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
2201 N. Dodge Street
P. O. Box 168
Iowa City, IA 52243

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Stephen Sireci
American Council on Education
GED Testing Service Suite 250
One Dupont Circle NW
Washington DC 20036

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550